

データ連携基盤の整備に向けた方針案
(目次)

資料1

1. 現状認識
2. 欧米の動向
3. 基本的な整備方針と目指すべき姿
4. 具体的ユースケース
5. データ連携基盤構築のために必要な事項と官民役割分担
 - (1) 技術的事項
 - ① 協調領域として実装すべき機能と研究開発課題
 - ② 競争領域として開発・サービス提供が期待される機能
 - ③ 特に留意すべき事項
 - (2) ルール・制度的事項
 - ① データ提供インセンティブ
 - ② データ利用権限、利活用を阻害しない仕組み
 - ③ データ品質の基準策定
 - ④ データ提供者、利用者の評価
 - ⑤ エコシステムの形成
 - (3) 推進体制
 - ① 共通語彙
 - ② データカタログ（メタデータ）
 - ③ A P I
 - ④ P D C Aサイクルによる段階的整備
6. 実現に向けたロードマップ

(参考) 各分野におけるデータ基盤の整備状況

データ連携基盤の整備に向けた方針案

1. 現状認識

Society 5.0は、サイバー空間とフィジカル空間の高度な融合により、全ての産業の生産性を飛躍的に向上させ、様々な社会課題を解決する社会である。Society 5.0を実現するためのカギは、国、自治体、民間などで散在するデータを連携させ、ビッグデータとして扱い、分野・組織を超えたデータ活用とサービス提供を可能とすることである。

現在、SIPなどの研究開発プロジェクトにより、農業、防災、自動走行などの分野において、分野内でのデータ連携の取組が進められ、成果をあげつつあるが、今後、分野をまたいだデータの連携を実現させるために、十分なサイバーセキュリティ対策のもと、分野間の「データ連携基盤」の構築を進めることが急務である。

2. 欧米の動向

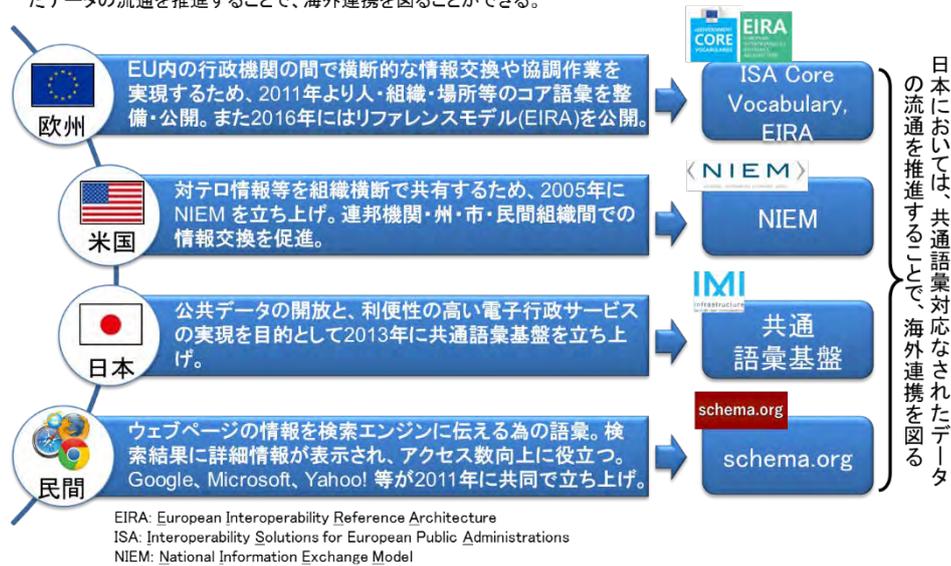
米国では2005年にNIEM (National Information Exchange Model) が、欧州では2011年にSEMIC (Semantic Interoperability Community) が稼働し、また、欧州では官民連携によるFIWAREが稼働するなど、データ連携の仕組みが存在している。データ連携は、組織横断的なデータの活用、意思決定やAI等の応用活用のために必須の基盤であることから、各国とも、データ連携基盤を将来に向けた中核政策として位置づけられている。

これまでも日本の共通語彙基盤であるIMI (Infrastructure for Multilayer Interoperability)、米国NIEM、欧州SEMICは情報交換を行ってきたところであり、今後、データ連携基盤の構築にあたり、欧米との一層の連携拡大が必要である。

EUでは、一般データ保護規則 (GDPR: General Data Protection Regulation) が2018年5月から適用されるため、個人情報保有する個人や企業、公的機関を含むすべての団体に、欧州経済領域内で取得したデータを域外に持ち出すには、十分な保護体制の整備を求められることになっており、民間企業はビジネス継続のため、対応が迫られている。

その中で、米国のMyDataイニシアティブ、英国のmidata、仏国のMesInfos等、データポータビリティを優先し、官民が保有するデータを再利用しやすい形で本人に還元し、本人関与の下でのデータ活用を拡大する施策が、各国において広がりつつある。

- 欧州や米国は日本に先行して語彙基盤の構築に着手。(欧州ISA / 米国NIEM)。2014年からは、EU、米国、日本が参加する各国語彙の連携会議が開始。民間では検索エンジン提供者の連合による schema.org が広く普及。
- 各語彙体系の相互運用性の確保を目指し、現在、活動中であることから、日本においては、共通語彙対応がなされたデータの流通を推進することで、海外連携を図ることができる。

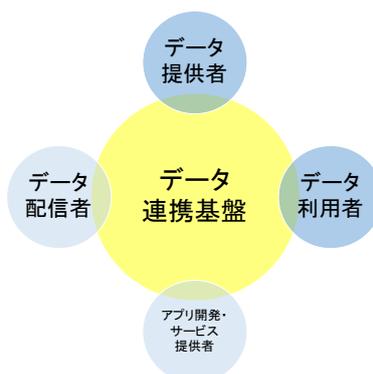


3. 基本的な整備方針と目指すべき姿

CSTI及びIT総合戦略室が司令塔機能を発揮し、関係各省や民間協議会等との連携のもと、あらゆるデータをワンストップで入手できるプラットフォーム「データ連携基盤」の構築を早急に進める。

データ連携基盤は、データそのものをストレージ（保有）せず、データカタログ（メタデータ）を用いて、欲しい産学官が保有するデータがどこにあるかを検索でき、標準API（Application Programming Interface：ネットワーク経由でデータ提供を行うアプリケーションインターフェース）を介して様々な分野のデータがワンストップで入手可能な分散・協調型プラットフォームである。

データ提供者は、データ連携基盤またはデータ配信者へデータを提供する。データ利用者は、データ連携基盤またはデータ配信者からデータそのものを入手、または、アプリ開発・サービス提供者を通じてサービス提供を受けることができる。



データ連携基盤の参画者

分野共通のコア語彙、分野共通で扱うドメイン語彙、ドメイン固有の応用語彙やデータ構造を整備し、分野横断でのデータのインターオペラビリティ（相互運用性）を確保する。なお、データ活用促進の観点から、個人情報保護やプライバシーに配慮しつつも、過剰な保護により利活用を阻害しないようにバランスをとり、これまでの発想や利害関係にとらわれない新規サービスを生み出せる制度・ルールを構築する。

これにより、データ連携基盤を活用した官民によるアプリケーションの創出等の様々なサービスへの展開、AIのための教師データとしての活用などが期待できる。また、データ流通推進協議会等の民間協議会により、個人データの市場となるPDS（Personal Data Store）や、情報銀行などの運用・技術基準の検討が進められているところであり、これら民間協議会と連携していくことにより、データ連携基盤が、PDSや情報銀行等と

データ利用者を仲介する基盤としての役割をも担うことが将来的に期待できる。

また、欧米との連携にあたっては、オープン・クローズ戦略などの知財戦略も意識しながら、語彙の相互互換、APIによる相互接続など、データ連携基盤の国際的なインターオペラビリティ（相互運用性）の実現に向けた連携を検討していくことが必要である。



データ連携基盤イメージ

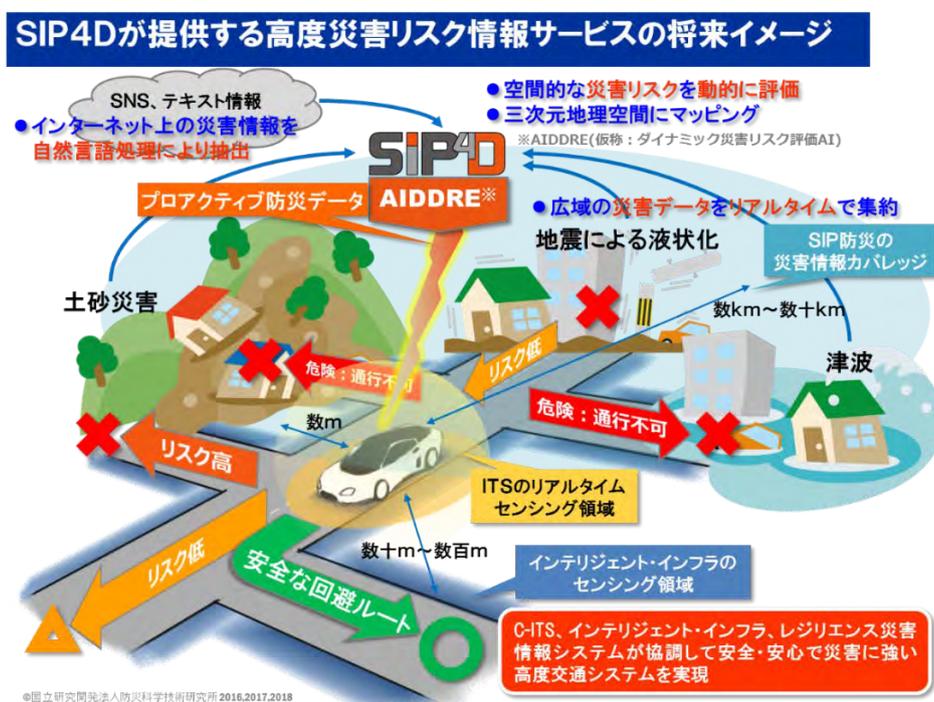
(参考) 新たな経済政策パッケージ 2017年12月8日閣議決定

国・自治体の各行政機関や企業等の民間機関の間で散在するデータをすべて連携することを目指し、「横断的分野」(位置、時間等)と「固有分野」(農業、インフラ等)双方について、データ標準や共通語彙基盤(IMI)の横断的なデータ活用を推進するための基盤を3年以内に整備することとし、そのためのシステム開発を開始する。並行して米国、欧州のデータ連携基盤とのデータ連携を検討し、日米欧10億人のデータ共通市場を創設することを目指す。

4. 具体的ユースケース

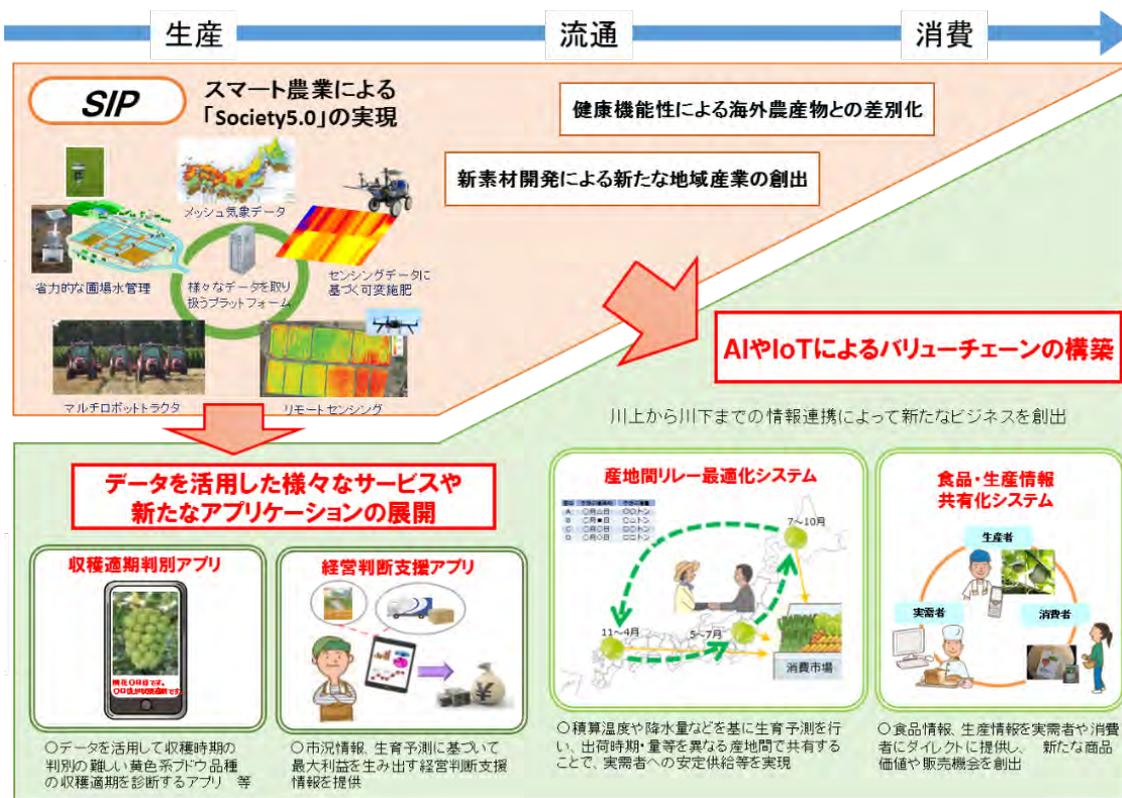
(1) 防災分野×交通分野

防災分野で基盤構築が進められているSIP4D (Sharing Information Platform for Disaster management) からの広域におけるリアルタイムな災害情報データを三次元地理空間情報などと分野横断でのデータ連携することにより、災害時の津波・洪水による浸水やインフラ被災状況等を踏まえた、安全な避難ルートの提供・避難誘導など、様々な連携の可能性が広がり、新たな価値創造が期待される。



(2) 農業分野×流通・消費分野 (スマートフードチェーン)

農業分野で基盤構築が進められている農業データ連携基盤からの食品情報、生産情報を実需者や消費者にダイレクトに提供することで、新たな商品価値や販売機会を創出することが可能となる。また、消費者のニーズと、地域毎の出荷時期や量を産地間で共有することで、作物の効率的な流通や安定提供が可能となる。

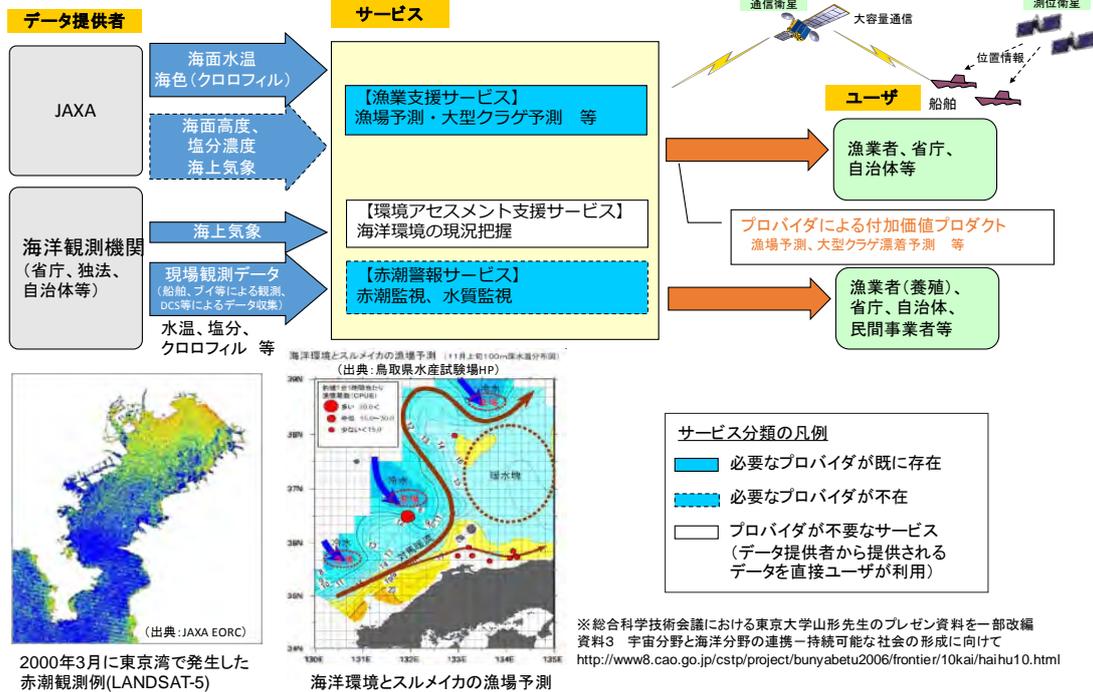


第 12 回重要課題専門調査会資料より

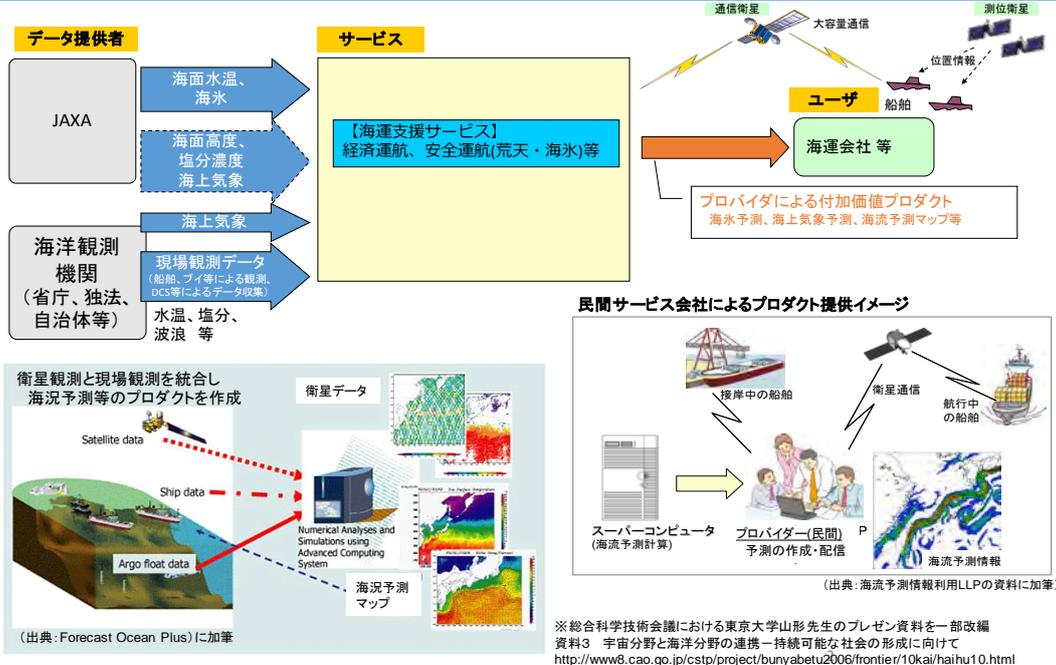
(3) 海洋分野 × 宇宙分野

航行船による現場観測データと、人工衛星による衛星観測データを組み合わせることで互いの時間・空間分解能を補完し、統合的な海洋観測網を形成できる。これらの観測網から、海水温、海色、海上風向・風速、海面高度、海流、高分解能画像等を活用することで、海流予測や船の位置、経路情報も一括してGIS (Geographic Information System) 上に紐づけることが可能になるため、水産業における漁場推定や赤潮監視としての利用や、海運業における船の安全や経済的な運航への利用が期待される。

水産支援

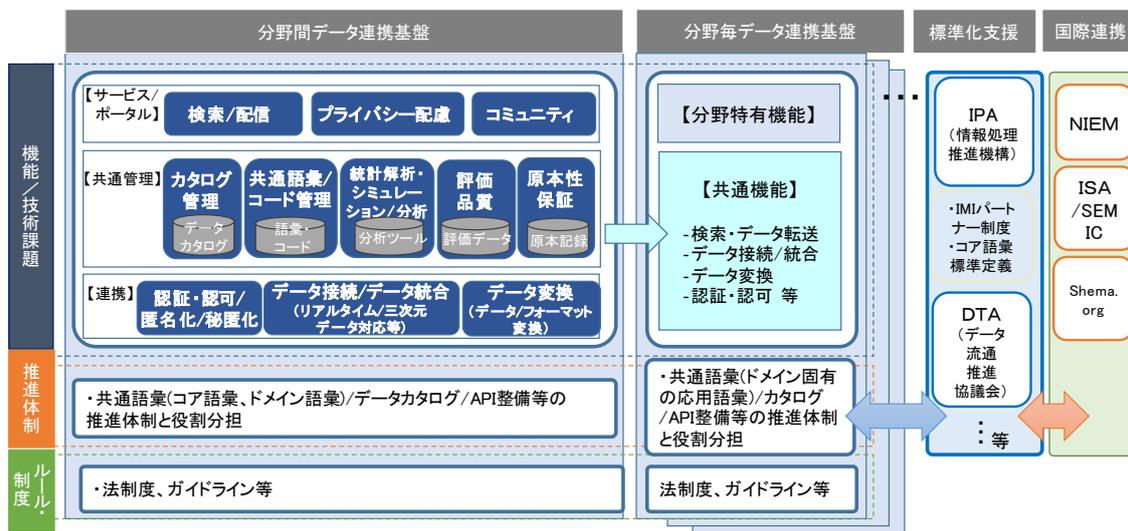


海運支援



5. データ連携基盤構築のために必要な事項と官民役割分担

データ連携基盤の構築にあたり、技術的事項、ルール・制度的事項及び推進体制について、以下、実施方針を示す。



データ連携基盤の全体イメージ

#	区分	機能	機能概要
1	サービス/ポータル	検索/配信	利用者がデータを利用し易くサポートする機能
2		プライバシー配慮	あいまいなキーワードにて検索し、複数分野のデータを集めて配信 IoTデータへの対応も含む
3		コミュニティ	プライバシーに配慮し、オプトイン・オプトアウト・データ利用目的の追加等の利用者・提供者とのやり取りの場を提供・記録
4		コミュニティ	利用者同士の意見交換・イノベーション協創(共同研究等)等の場を提供
5	共通管理		データ検索や連携のためにデータ連携基盤が管理する機能
6	共通管理	データカタログ管理	データカタログ(データの名称、作成者、作成日など)を登録、更新
7		共通語彙/コード管理	データを共通項目名等に揃えるための語彙/コード情報を登録、保管
8		統計解析・シミュレーション/分析	分野共通のシミュレーションや相関等分析(EBPM向け相関分析含む)
9		利用者・提供者評価/データ品質基準	利用者・提供者の相互運用性、信用度等やデータの品質(完全性、正確性、有用性、最新性等)等の評価値を記録・更新。認証機能と連携しアクセスを制御
10		原本性保証	各分野で発生するデータの原本性を保証。データ流通品質を担保
11	連携		データ利用者の要求に応じて、データ提供者のデータを応答する機能
12	連携	認証・認可/匿名化/秘匿化	提供者、利用者の認証、データ匿名化、暗号化等
13		データ接続	データ提供者、利用者との接続を行う機能。(センサデータ等のリアルタイム接続、三次元データ接続も含む)
14		データ変換	データの単位、座標系、項目名等及びデータフォーマットを揃える機能

(1) 技術的事項

価値創造、利益の源泉になりうる機能やツールについては、「競争領域」として、個別の民間企業が独自の技術やアイデアを活かして開発すべきである。一方、データの流通や利活用の促進につながる基盤となる機能やツールについては、「協調領域」として、内閣府・各省連携のもと、戦略的イノベーション創造プログラム（S I P）、官民研究開発投資拡大プログラム（P R I S M）等のプログラムを活用し、官民協力のもと整備を推進する。

① 協調領域として実装すべき機能と研究開発課題

i) 実装すべき機能

（データカタログ管理機能）

国・自治体の各行政機関、民間企業、大学や研究機関に散在するデータのそれぞれがどこにあるのか、どんなデータがあるかをメタデータ（項目）に示し、データカタログとして整備する。これにより、データ連携基盤の利用者が欲しいデータを検索しアクセス可能とする。その際、I o T機器で測定される観測系データや、アンケートなどの調査系データなど各種分野の違いについても配慮し、表現できるメタデータの項目、形式を調査決定する。

（認証・認可機能）

営業機密データ、セキュリティに関わる機微な情報など、様々な特性を持つデータを適正に流通させるため、データ利用者とデータ提供者を認証し、あらかじめ設定されたデータの利用権限、契約関係に基づき、適切に認可、アクセス制御する機能を実装する。

（検索機能）

利用者の誰もが、欲しいデータを容易に探しだすことができるように、データカタログ管理機能と連携し、キーワード検索の機能をデータ連携基盤で提供する。また、利用者の利便性向上の観点から、自然言語処理技術や機械学習技術などを活用した高度な検索、リコメンド機能の開発・実装が期待される。

（データ接続/変換/統合機能）

汎用的に取り扱われるデータ形式（JSON、XML、CSV、GeoJSON、Shape、GML）を対象としたフォーマット変換、同種の表形式（xls等）データのデータ統合、地図データ（GISや写真画像）への地物情報の紐づけなどに

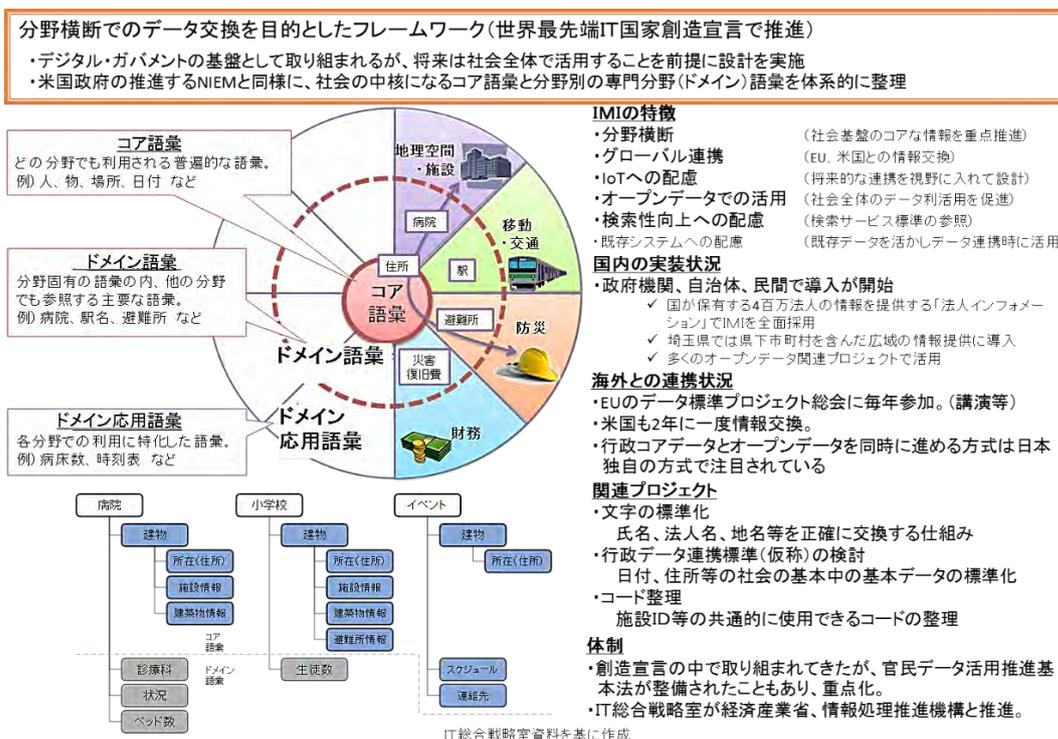
については、分野を問わず共通的に利用が想定され、協調領域として整備することを検討する。

(共通語彙・コード管理)

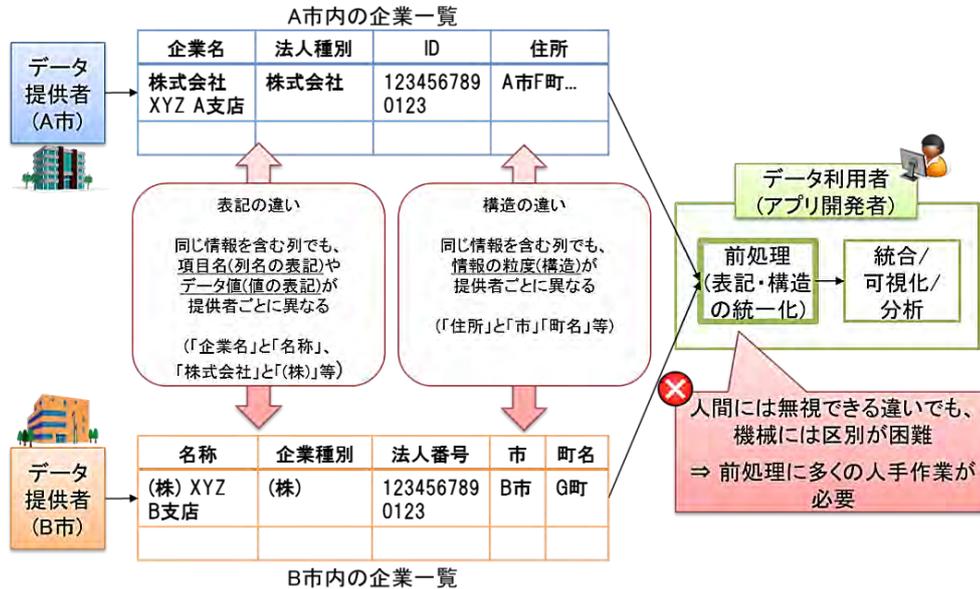
異なる分野間でデータを流通させ利活用するため、どの分野でも共通して用いられるコア語彙（例えば、人、物、住所、日付等）とコードについては、IT総合戦略室主導のもと、情報処理推進機構（IPA）の共通語彙基盤（IMI）の中で、今後益々増大するIoTデータも考慮した上で継続的に整備する。

各分野の担当省庁や関連する民間コンソーシアムは、各分野内で特定の課題やシステムで用いられるドメイン固有の応用語彙（例えば、病床数、時刻表など）を整備する。また、各分野の基礎的な用語であって分野を超えた利用もなされるドメイン語彙（例えば、病院名、駅名、避難所）を整備する。

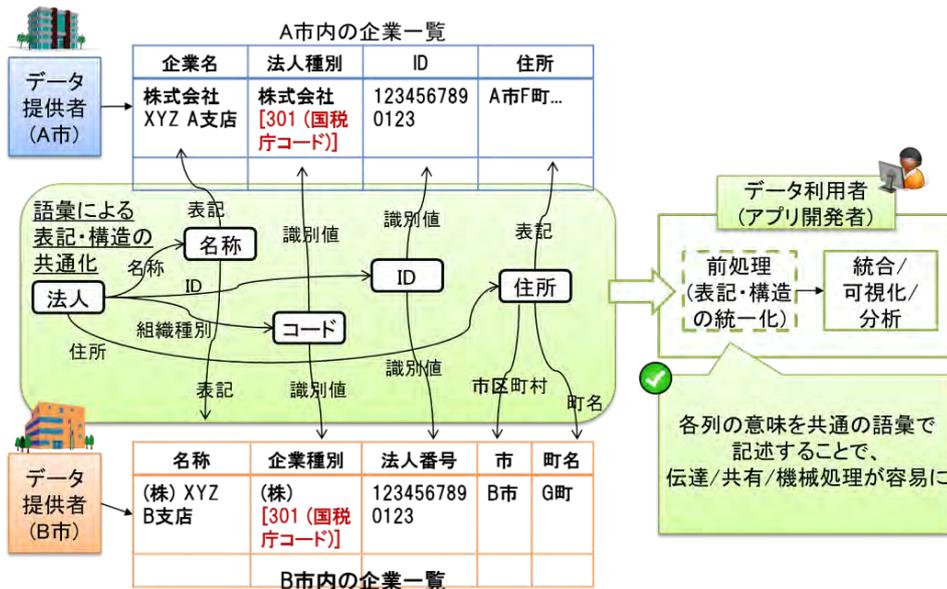
また、語彙の整備にあたっては、日米欧での相互運用性を確保するため、米、EUの行政主導で整備されているNIEM、SEMICに加えて、ウェブ上の検索のために民間主体で整備されてきた schema.org も参照し、相互に対応付ける。また、W3C、ITU、UN/CEFACT、ISO/IECの関連する語彙規定の動きも参照する。



- 現状、表記や構造が異なる複数のデータを統合/可視化/分析するには、データ利用者は、それらの表記や構造を統一するためのデータ修正に多くの時間/手間を要している。



- 共通語彙基盤の語彙やコードを用いることで、異なる表記や構造をもつ複数のデータに共通の表記や構造を与えることができ、データの共有/機械処理が容易になる。



（統計解析/分析機能）

国の証拠に基づく政策立案（EBPM: Evidence-Based Policy Making）や民間企業のマーケティング等で共通的に用いられる一般的な集計や相関分析等の統計処理機能については、データ連携基盤の魅力的なコンテンツになる可能性があり、データ提供者、利用者のデータ連携基盤利用のインセンティブに繋がるため、協調領域として整備することも検討する。

（プライバシー配慮）

プライバシーに配慮し、オプトイン、オプトアウト、利用目的の追加や更新など、利用者と提供者の間で必要なやりとりを行うためのインターフェースの提供とその情報の履歴や更新状態を記録し、アクセス制御と連動させる機能を実装する。

（コミュニティ）

アプリケーション開発者などのデータ利用者が、オープンソースソフトウェアとして、率先してデータ処理ツールや分析ツールなどを公開・更新する場（例えば、GitHub等のウェブサイト）をデータ連携基盤上に提供する。また、データ連携基盤運営者への改善要望の取り込みや、普及活動、広報のための情報発信元としての活用が期待される。

また、利害関係が一致するコンソーシアムやベンチャー企業に、クラウド環境を期間限定で提供することで新事業のスタートアップを支援することもデータ連携基盤の活性化のため重要である。

ii) 研究開発課題

データ連携基盤の実現にあたっては、データ提供者がメタデータをデータ連携基盤に登録する際の作業負担をできるだけ省き、また、提供者にもデータ連携基盤を利用するメリットが感じられる機能/ツールを開発・実装する。これらの機能/ツールについては、グローバルにオープンソース化することも検討する。

（データの機械可読性向上）

データ利用者が、データ連携基盤を介して収集・共有されたビッグデータをAI等により分析し、サービスに活用するためには、それぞれのデータがコンピュータで処理できるように機械可読性を確保された状態にする必要がある。

例えば、紙に手書きで保管されている台帳を効率よくデータ化するため

の支援ツールや、エクセル形式で保管されたデータのセルの結合解除、文字間のスペース除去、キャプションの削除等、前処理に関する作業を軽減するためのツールなど、汎用的で頻繁に発生する作業については、データ連携基盤において開発・実装することが求められる。

機械判読不可データを、機械判読可能なデータに変換するための処理例を以下に示す。

・キャプション、脚注、脚注番号の削除

表外のキャプション、脚注、及びセル内の脚注番号などを削除する。

表形式データの架装データサンプル(その1)

集計項目	平成23年 (1000円)	24年(1000 円)	差分 (1)
合計	55000	127788	▲1,232
あ	1900	1100	110
い	2000	2200	110
う	3000	3300	110
え	4000	4400	110
お	5000	3300	66
か	6000	2200	37
き	7000	1100	16
く	8000	5500	89
け	9000	9900	110
こ	10000	10000	100



集計項目	平成23年 (1000円)	24年(1000 円)	差分
合計	55000	127788	232
あ	1000	1100	110
い	2000	2200	110
う	3000	3300	110
え	4000	4400	110
お	5000	3300	66
か	6000	2200	37
き	7000	1100	16
く	8000	5500	89
け	9000	9900	110
こ	10000	10000	100

(※1)

・セル結合の解除

セルが結合されている場合は、セル結合を解除し、全てのセルに元の値をコピーする。

年度	期	A (円)	B (円)	C (円)
2005	上	0.01	0.01	0.01
	下	0.01	0.01	0.01
2006	上	0.01	0.01	0.01
	下	0.01	0.01	0.01
2007	上	0.01	0.01	0.01
	下	0.01	0.01	0.01
2008	上	0.01	0.01	0.01
	下	0.01	0.01	0.01
2009	上	0.01	0.01	0.01
	下	0.01	0.01	0.01
2010	上	0.01	0.01	0.01
	下	0.01	0.01	0.01



年度	期	A (円)	B (円)	C (円)
2005	上	0.01	0.01	0.01
2005	下	0.01	0.01	0.01
2006	上	0.01	0.01	0.01
2006	下	0.01	0.01	0.01
2007	上	0.01	0.01	0.01
2007	下	0.01	0.01	0.01
2008	上	0.01	0.01	0.01
2008	下	0.01	0.01	0.01
2009	上	0.01	0.01	0.01
2009	下	0.01	0.01	0.01
2010	上	0.01	0.01	0.01
2010	下	0.01	0.01	0.01

(※1)

※1) 各府省情報化統合責任者(CIO)連絡会議決定(平成25年6月25日)「二次利用促進のための府省のデータ公開に関する基本的考え方(ガイドライン)」より
<http://www.kantei.go.jp/jp/singi/it2/cio/dai52/kihon.pdf>

(参考) 機械可読性の乏しいデータに対するデータ化支援ツールの例

(メタデータ作成支援ツール)

データ提供者が自らのデータをデータ連携基盤に接続する際に、データ連携基盤に登録するデータ形式であるメタデータを作成する必要がある。各分野において既存のデータ基盤が存在する場合、その既存のメタデータからデータ連携基盤に登録すべきメタデータへの自動変換のためのツールや、メタデータの作成に不慣れな作業員にも、容易に、効率的に作業が行えるよう、入力支援画面などをデータ連携基盤で開発し実装することは、より多くの分野から、データ連携基盤にデータを提供してもらうために重要である。

(AIを実装した類似語推定機能)

異なる用語で記述された同義語に対して、類似語として推定し提示する機能など、共通語彙基盤の整備だけではカバーしきれない語彙への対応は、

AI 技術を活用することで、技術的に解決される可能性がある。これは、データを統合する際や、検索する際に利用者の利便性を各段に向上させる機能であり、データ連携基盤で研究開発し、実装することが期待される。

(原本性保証)

データ流通が加速し、例えばサプライチェーンでデータの利用・再販が進む際、データ利用者が安心してデータ利用するために、データ品質担保が必要になるため、データの原本性を保証するための機能を官民の適切な役割分担のもとで研究開発を行い、データ連携基盤に実装することを検討する。

② 競争領域として開発・サービス提供が期待される機能

(シミュレーション/分析機能)

データ連携基盤の効果的な活用例として、データ連携基盤を介して集約された複数のデータをビッグデータ化し、AI を活用した高度なシミュレーションを行うこと等が考えられる。こうした各社独自技術の要素が強く、利益の源泉にもなる機能については、競争領域として開発されることが望ましい。

(高度なデータ接続/データ統合機能)

IoT センサーから日々生成される複数のリアルタイムデータを高速に統合する機能をはじめとする革新スピードが速い技術は、国が主導して開発・実装するよりも、競争原理が働き最新技術の導入が早い民間企業に任せることが望ましい。

(匿名化/秘匿化機能)

個人情報を含むデータの匿名化/秘匿化については、扱うデータやその利用目的によって、どのようなレベルで行うべきかが異なるため、各分野のルールに従って、各分野のデータ基盤で取り扱うべきと考えられる。

また、第三者提供に関する本人同意に基づき、情報銀行や PDS (Personal Data Store) に提供されたデータに関しては、情報銀行や PDS で決められた利用権限をデータ連携基盤の中に継承することで、データを流通させることも検討する。

③ 特に留意すべき事項

i) 分野毎データ基盤以外からの接続

内閣府のSIPやPRISMをはじめとする国や自治体で進められている各プログラムで構築、また、今後構築される分野毎データ基盤は、この分野横断のデータ連携基盤との相互運用性を確保すべきである。

さらに、各分野のデータ基盤に属さない国、自治体、大学、民間企業が分散して保有するデータについては、分野毎データ基盤を経由しなくても、直接、データ連携基盤へのデータの提供、データカタログの登録が可能となるよう、共通語彙基盤、API、メタデータの仕様等を広く公開し、連携促進を図る。

ii) 民間データセンターの活用

防災、農業、材料、地球環境、海洋分野等では、すでに国立研究開発のリーダーシップのもと、分野毎のデータ基盤構築が推進されている。一方、データセンターについては、その維持管理やサイバーセキュリティ対策に必要な人的リソースの継続的な確保が懸念される。また、最先端のICT技術の導入の観点からも、分野毎のデータ基盤、分野横断のデータ連携基盤は競争原理が働く民間企業が運営する最新のデータセンターを活用することが望ましい。

iii) 相互運用性の確保

既に様々なデータ、システムが稼働しているところであり、データフォーマット、語彙、メタデータ、API等を全面的に標準化ありきで進めるのではなく、相互運用性を優先し、変換機能の実装など、技術的な解決手段によって、合理的に実現するべきである。

民間等の独自のデータ提供サービス構築等を阻害することなく、できることから連携を進めていくことが重要である。また、ICT技術の進展に合わせて将来的な変更・アップグレードを前提とした考えのもとで、柔軟なシステム構築を目指すべきである。

iv) サイバーセキュリティの確保

IoTで全てのヒトとモノがつながる Society 5.0では、サイバー攻撃の起点が増大するとともに、複雑につながるサプライチェーンを通じてサイバーリスクの範囲が拡大する。また、サイバー攻撃による影響がフィジカル空間にまで達するリスクがある。データ流通市場の活性化が進み、大量のデータがグローバルサプライチェーンにおいて連携し、データの利用・

再販が進むことを想定すると、ハイレベルなサイバーセキュリティ対策を備えたデータ連携基盤を構築することが重要である。

(2) ルール・制度的事項

データの流通や保護に関する制度や知財戦略、GDPRを始めとする諸外国の動向等も踏まえ、データ連携基盤の利活用が促進されるルールや仕組みの整備を進める。

① データ提供インセンティブ

データ連携基盤をより効果的で価値のある基盤として整備するためには、新たなサービスやイノベーションの源泉となる価値のあるデータをいかに多くデータ連携基盤で流通できるかが鍵となる。

しかしながら、分野毎のデータ基盤運営者やデータ提供者にとって、経営資源ともなるデータの提供や作業負担が発生するメタデータの登録は、相応のメリットがないと協力が得られにくい。

このため、データ提供者への対価（課金制度や提供者の評価への反映等）の還元の仕組みや、公的資金による研究開発成果の一次データの公開推奨（例えば、機械可読性のあるデータで、〇年以内）等のガイドライン化などを検討することが必要である。まずは、公的データを率先して登録し、基盤としての魅力を向上させることが必要である。

また、分野間連携の具体的な姿を想定した上で、象徴的な案件として実証試験をすることで良好事例を作り、外部へメリットを発信することが重要である。（4. 具体的ユースケース 参照）

自治体提供のオープンデータ等、複数機関が同種のデータを提供する場合には、データの相互運用性を確保するために、データ提供者向けにデータ変換機能などのツールで支援することや、データの登録・チェックを代行する業者を利用できる仕組みを整備すること等が必要である。

また、データを提供することによって、個人情報漏えいや事故などの責任問題が発生しないかを危惧し、データ提供者がデータを出すことに対して感じるハードルを下げることも重要である。そのためには、例えば一定条件下での免責を定めるなどの利活用ルールの検討等が考えられる。

② データ利用権限、利活用を阻害しない仕組み

データ連携基盤においては、営業秘密データ（民間企業が持つ価値の高いデータ等）、機微なデータ（国家安全保障に関わるもの等）、個人情報を含むデータ、オープンデータ等の様々なデータを対象に分野横断で流通す

ることを想定し、それぞれのデータの特質、利用権限に応じたアクセス権限の付与が必要である。特に、個人情報保護、プライバシーへの配慮をしつつも、過度な保護による利活用の阻害につながらないように、データ連携基盤でのデータ利用に関するルールを整理することが必要である。

たとえば、

- ・ オープンデータについては利用者登録も必要とせず、フルオープンとする。
- ・ 機微なデータについては、利用者を登録した上で、適切なアクセス制御を付与する。
- ・ 営業秘密データについては、データ連携基盤とは別に、契約プロセスを経た上で、提供者と利用者の双方のアクセス権を設定する。
- ・ 個人情報を含むデータに関しては、提供者側で匿名加工済みのデータをデータ連携基盤で扱う。または、第三者提供に関する本人同意に基づき、情報銀行やPDS（Personal Data Store）に提供されたデータに関しては、情報銀行やPDSで決められた利用権限をデータ連携基盤の中に継承することで、データを流通させることも検討する。

また、匿名加工済みデータの組合せにおける個人特定に対する責任については、免責等のルール決めが必要である。

データの提供と利活用の促進の観点から、データ利用規約についてはデータ連携基盤の中では可能な限り類型化し、提供者が決めた規約に配慮しつつも、類型化された規約に合わせるように要請する。また、データ利用にあたり、提供者名のクレジット表示が求められる場合があるが、要否や表示方法のルールについては検討が必要となる。

データ連携基盤上の利用権限に係るルールの検討にあたっては、ビッグデータを活用したサービス等のための著作物の利用に関する権利制限規定（著作権法改正案）等の関連制度整備を踏まえることが重要である。

③ データ品質の基準策定

データ利用者が、アプリ開発やサービスを提供する際に、データの完全性、正確性、有用性、最新性等の品質を正確に知る必要があり、データそのもの入手する前に、データカタログのメタデータ等でデータ品質を知りたいというニーズがある。今後、基盤として評価の基準づくりに取り組み、それぞれのデータ品質を統一した基準のもとで表現することが望ましい。

その上でデータ利用者は、求める品質のデータを適切に選択することが

可能となるため、活用に関してはデータ利用者側の市場原理に任せることができる。

このように、データ提供者にとっては、メタデータ作成の作業負荷が大きいものの、利用者にとっては適切なデータを発見するのに役立つため、データ連携基盤では、必須項目（作成日、作成組織、最終更新日、ライセンス等）に加えて、データ品質を表現する任意項目を設けるとともに、メタデータ作成を支援する仕組みや体制を整備することが必要である。

④ データ提供者・利用者の評価

データの提供/利用を安心して行うためには、データ提供者・利用者の信用度等の評価基準や評価方法を整備し、データ提供者・利用者が、データを提供/利用する対象を選定するための指標を提供する必要がある。信用度の評価は、他のセキュリティ基盤と連携し、アクセス制御も合わせて行うことは、安心・安全なデータ流通のために重要である。

⑤ エコシステムの形成

国の資金に頼らず、持続的にデータ連携基盤を自立運用するためには、データ連携基盤の運用・管理・保守等を考慮したエコシステムの形成が不可欠となる。

データ連携基盤のビジョンは、官民の散在するデータに加え、我が国の強みであるセンサー／デバイスから集められるリアルタイムデータをも活用し、新たなサービスの提供や、AIの学習データとしてのビッグデータを提供するサービスプラットフォームとして機能させることである。

そのため、民業圧迫に配慮しつつも、データの変換や分析、クラウド環境提供等の基盤としての付加価値を生む機能についても、実装を前向きに検討する必要がある。

また、サービス創出の活性化やデータ活用の更なる利便性向上のため、利用者・提供者が基盤を通じてコミュニケーションを図りやすい場を提供することで、データ連携基盤の自律的な成長を図る。

データ連携基盤の開発にあたっては、内閣府、IT総合戦略室が司令塔機能を発揮し、官民でマルチステークホルダーの分野横断の調整・合意形成の場を提供し、継続的な成長を促す。その上で、将来的な運営主体に関しては、公的データの管理は国等が担う一方、それ以外の基盤運営等については、民間へ拡張すること等を検討する必要がある。

(3) 推進体制

データ連携基盤の整備にあたっては、内閣府とIT総合戦略室が司令塔機能を発揮し、SIP、PRISMを活用した官民連携体制を進めることを基本とする。分野横断的にデータを連携させるために、共通語彙、メタデータ、APIの整備が特に重要な要素となる。

分野毎のデータ基盤においては、分野間連携に必要となる語彙やメタデータ等の整備は十分に進んでおらず、各分野の責任主体は、分野毎にデータ連携基盤に接続するデータセット、語彙、コード、APIについて、整理する必要がある。

分野間データ連携を実現する上で、国や自治体がリードすべき行政データにおける共通語彙基盤の整備やデータの構造化については、専門知識が必要なため、IT総合戦略室と情報処理推進機構（IPA）を主体として、民間団体等の協力も得つつ、各省庁を支援する体制が想定される。一方、産業分野が主体のものづくりやIoTデータのデータ構造、メタデータ項目の検討については、民間企業のコンソーシアム等で合意形成をし、主体的に進めることが想定される。

① 共通語彙

分野毎のドメイン固有の応用語彙については、分野毎のデータ基盤構築をリードする国研や関係省庁が、分野内の関係者との調整を主体的に実施する。その際に、ウェブ上の検索精度向上のために民間中心で策定されている schema.org は対象分野が広く、IoTのデータモデルへも再利用が進められているため、スマートシティや観光等の分野では参照されることが望ましい。

分野共通で用いられるコア語彙については、IPAが相互運用性の拡充を目指す賛同企業の集まりであるIMIパートナーとも連携し共通語彙基盤の取組が既になされており、これを継続的に推進する。

分野横断で相互に連携するためのドメイン語彙については、関係府省が国研などの分野毎データ基盤の責任主体を集め、IPAや民間コンソーシアム等の支援のもとで調整する。

② データカタログ（メタデータ）

IT総合戦略室を主体として、これまでのオープンデータの取組みや、今後、関係各省庁・民間団体へのヒアリング等から得る情報を元に、先行分野を定め、行政データに関するメタデータの項目を決定する。また、主に民間で進んでいるIoTデータへの対応に向け、データ流通市場の健全

な成長を目的とした民間団体（データ流通推進協議会等）とも連携する。

③ A P I

データ提供者ごとにデータ要求・取得インターフェースが異なると、各 A P I に合わせた整備が必要となる。効率的なデータの利活用のため、A P I の正規化/共通化が必要である。

2017年8月にIT総合戦略室から、「A P I 設計・運用実践 ガイドブック（β版）」が公開されており、これを参照することで、共通的な A P I 設計が可能となる。

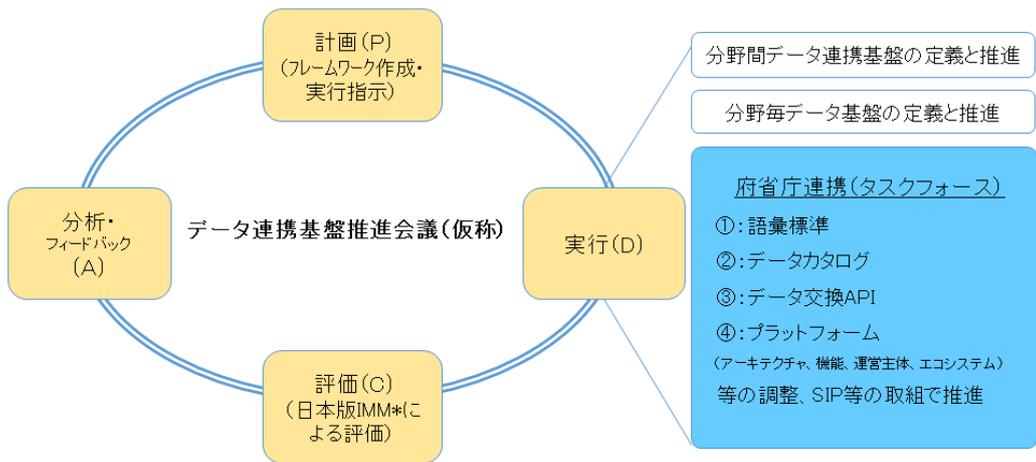
しかしながら、既に構築中または運用中の分野毎のデータ基盤やシステムでは、A P I の仕様が異なるため、今後、分野間、システム間の相互運用性を確保するためには、これらの A P I を正規化/共通化することが望まれる。そのため、IT総合戦略室が N I E M、S E M I C、F I W A R E などの欧州の取組や、民間主体のコンソーシアム等とも連携しながら、国際標準（O M A / N G S I 等）を参考・拡張し、A P I の仕様策定および実装を行う。

④ P D C A サイクルによる段階的整備

メタデータの整備、データの構造化、語彙の整備などの利用者目線での段階的な整備を進め、特定分野・エリア（自治体等）での実証を行いつつ、P D C A サイクルを回しながら基盤の構築を進める。

データ連携基盤の進捗や具体的なユースケースの共有など分野毎のデータ基盤との連携を進めるとともに、分野毎データ基盤の成功事例や実運用の経験を取り込むため、定期的な連携会議を開催する。また、将来的な変更・アップグレードを前提に、必要に応じて有識者によるアーキテクチャレビューを行う。

- データ連携基盤推進会議(仮称)を設置し、機能要件、アーキテクチャに加え、データ提供、利用に伴うルール、制度、基盤の運営を含め継続的に議論
- 関係府省庁との調整のため、タスクフォースを実施



*IMM: Interoperability Maturity Model
 欧州の取組を参考に、データの相互運用性成熟度モデル (IMM) を策定し、各サービスの相互接続性(相互データ提供がどの程度可能か)に関するレーティング、改善活動等を実施

データ連携基盤の整備のためのPDCAサイクル

6. 実現へ向けたロードマップ

研究項目	2018年	2019年	2020年	2021年	2022年
マイルストーン			△2020オリバラ		
	【フェーズ1】 2020オリバラでのデモも想定した、試行に向けた研究、環境構築、有効性検証等		【フェーズ2】 2023年度以降の本番運用を想定した研究、環境構築、有効性検証等		
技術研究・開発 (協調領域)	マスタプラン作成 語彙・コード・カタログ管理、検索、コミュニティ機能他		マスタプラン見直し 有効性検証		
		利用者・提供者評価、検索、等			有効性検証
ルール・制度的 事項の検討・整備	データ利用権限、データ品質 他 の 検 討 ・ 施 策 整 備				エコシステム検証
	エコシステムの検討				
推進体制の 整備と実行	共通語彙、メタデータ、API検 討体制整備	共通語彙、メタデータ、API検討体制によるコンテンツ整備			
			分野毎データ連携基盤との連携		